Solving Research Problems: Analyze Mouse Embryonic Stem Cell Transcriptional Profiles

Molecular Analysis Tools Knowledge Center

Molecular analysis tools provide powerful analytic capabilities for genomic data interpretation. Four key tools are supported by the Molecular Analysis Tools Knowledge Center: geWorkbench, which provides an innovative, open-source software platform for genomic data integration while bringing together analysis and visualization tools for gene expression, sequences, pathways and other biomedical data; GenePattern, which provides bioinformatics tools for gene expression, proteomic and SNP analysis; caArray, a system that supports the management and exchange of array data and annotations and caIntegrator, a novel translational informatics platform that allows researchers and bioinformaticians to access and analyze clinical and experimental data across multiple clinical trials and studies.

The Problem: Understanding Mouse Embryonic Stem Cell Transcriptional Profiles

Dr. Bradley Merrill of the University of Illinois at Chicago and his lab recently performed their first gene expression array experiment comparing a mutant mouse embryonic stem cell line to a non-mutant control line, with the help of his university's genomics core facility. The genomics core facility normalized and annotated the raw data, performed gene ontology and pathway analyses, and returned the results to Dr. Merrill. "While these analyses were fine, I felt that we were being too superficial," explained Dr. Merrill. He began to look for his own solution to perform more complex analyses.

The Solution: GenePattern

Dr. Merrill evaluated several computer applications, typically spending a week with each software package to see if that application provided the data analysis he needed. He said, "I abandoned using the other software soon after downloading GenePattern. It took me about two days to learn how to use the software for the most basic analyses like RMA normalization of expression levels and Comparative Marker Selection, which identified expression defects in our mutant stem cells." He noted that a number of the genes deregulated in his mutant cells were previously reported to be important for stem cell self-renewal, and were regulated by stem cell factors Nanog and Oct4. Based on the phenotype of the mutant stem cells and the genes that were deregulated in the mutants, he hypothesized that the mutant was affecting stem cells by regulating the same genes as Nanog and Oct4.

He tested this hypothesis by comparing gene expression in the mutants with gene expression changes caused by reduction of Nanog or Oct4. The Nanog and Oct4 data was published by another group and was available online through NCBI. He used GenePattern to normalize raw data from the other group with the same parameters as he had done with his own data. This allowed him to examine whether the changes caused by Oct4 or Nanog were similar to the changes caused in his mutant cell line. He used the Venn diagram and Hierarchical clustering capabilities within GenePattern to demonstrate the similarity in effects caused by the different factors, and then used Spearman rank order and Pearson correlation tests (not part of GenePattern) to demonstrate a strong statistical significance of the similarities. He said, "These analyses were critical for our understanding of how our mutant affects stem cells and were important for the publication of a manuscript from my lab: Yi F, Pereira L, Merrill BJ. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells*. 2008 Aug;26(8):1951-60.)"

The Benefits:

GenePattern remains important for Dr. Merrill's research, enabling him to "examine biological problems using a whole new approach." He found that GenePattern allows him to test hypotheses directly, and in less time than it would take him to communicate with a bioinformatics specialist. He continues to use GenePattern frequently, usually to compare signatures of gene expression from his results with new datasets. He added, "The ability to normalize raw data gives me confidence that I can compare gene expression in an unbiased manner. Honestly, I feel that it is quite empowering."

Key Contributors:

The Broad Institute of MIT and Harvard
Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago

For More Information:

Molecular Analysis Tools Knowledge Center: *https://cabig-kc.nci.nih.gov/Molecular/KC/index.php/Main_Page*

caBIG® Knowledge Center